

an underscore bar and six digits:

NT_123456	constructed genomic contigs
NM_123456	mRNAs
NP_123456	proteins
NC_123456	chromosomes

Note: compare accession number with Sequence Identifiers such as Version and GI for nucleotide sequences, and ProteinID and GI for amino acid sequences.

Entrez Search Field: Accession [ACCN]
Search Tip: The letters in the accession number can be written in upper or lower case. RefSeq accessions must contain an underscore bar between the letters and the numbers, e.g., NM_002111.

VERSION

A nucleotide sequence identification number that represents a single, specific sequence in the GenBank database. This identification number uses the accession.version format implemented by GenBank/EMBL/DDBJ in February 1999. ↑

If there is any change to the sequence data (even a single base) the version number will be increased, e.g., U12345.1 --> U12345.2, but the accession portion will remain stable.

The accession.version system of sequence identifiers runs parallel to the GI number system. That is, when any change is made to a sequence, it receives a new GI number AND an increase to its version number.

For more information, see section 1.3.2 of the GenBank 111.0 release notes, and section 3.4.7 of the current GenBank release notes.

A Sequence Revision History tool is available to track the various gi numbers, version numbers, and update dates for sequences that appeared in a specific GenBank record (more information and example).

Entrez Search Field: Can use either Accession [ACCN] or UID

• GI

"GenInfo Identifier" sequence identification ↑

number, in this case, for the nucleotide sequence. If a sequence changes in any way, a new GI number will be assigned.

A separate GI number is also assigned to each protein translation within a nucleotide sequence record, and a new GI is assigned if the protein translation changes in any way (see below).

GI sequence identifiers run parallel to the new accession.version system of sequence identifiers. For more information, see the description of Version, above, and section 3.4.7 of the current GenBank release notes.

Entrez Search Field: UID

KEYWORDS

Word or phrase describing the sequence. If no keywords are included in the entry, the field contains only a period. ↑

The Keyword field is present in sequence records primarily for historical reasons, and is not based on a controlled vocabulary. Keywords are generally present in older records. They are not included in newer records unless (1) they are not redundant with any feature, qualifier, or other information present in the record, or (2) the submitter specifically asks for them to be added, and (1) is true, or (3) the sequence needs to be tagged as an EST, STS, GSS or HTG.

Entrez Search Field: Keyword [KYWD]

Search Tip: Since keywords are not present in many records, it is best not to search that field. Instead, search All Fields [ALL], the Text Word [WORD] field, or the Title Word [TITL] field, for progressively narrower retrieval.

SOURCE

Free-format information including an abbreviated form of the organism name, sometimes followed by a molecule type. (See section 3.4.10 of the GenBank release notes for more info.) ↑

Entrez Search Field: Organism [ORGN]

Search Tip: For some organisms that have well established common names, such as baker's yeast, mouse, and human, a search for the common name will yield the same results as a search for the scientific name. E.g., a search for "baker's